Fine-Scale Mapping of the FGFR2 Breast Cancer Risk Locus: Putative Functional Variants Differentially Bind FOXA1 and E2F1

Kerstin B. Meyer,[1,*] Martin O'Reilly,[1] Kyriaki Michailidou,[2] Saskia Carlebur,[1] Stacey L. Edwards,[3,4] Juliet D. French,[3,4] Radhika Prathalingham,[1] Joe Dennis,[2] Manjeet K. Bolla,[2] Qin Wang,[2] Ines de Santiago,[1] John L. Hopper,[5] Helen Tsimiklis,[6] Carmel Apicella,[5] Melissa C. Southey,[6] Marjanka K. Schmidt,[7] Annegien Broeks,[7] Laura J. Van 't Veer,[7] Frans B. Hogervorst,[7] Kenneth Muir,[8] Artitaya Lophatananon,[9] Sarah Stewart-Brown,[9] Pornthep Siriwanarangsan,[10] Peter A. Fasching,[11,12] Michael P. Lux,[11] Arif B. Ekici,[12] Matthias W. Beckmann,[12] Julian Peto,[13] Isabel dos Santos Silva,[13] Olivia Fletcher,[14] Nichola Johnson,[14] Elinor J. Sawyer,[15] Ian Tomlinson,[16] Michael J. Kerin,[17] Nicola Miller,[17] Federick Marme,[18,19] Andreas Schneeweiss,[18,19] Christof Sohn,[18] Barbara Burwinkel,[18,20] Pascal Guénel,[21] Thérèse Truong,[21] Pierre Laurent-Puig,[22] Florence Menegaux,[21] Stig E. Bojesen,[23] Børge G. Nordestgaard,[23] Sune F. Nielsen,[23] Henrik Flyger,[24] Roger L. Milne,[25] M. Pilar Zamora,[26] Jose I. Arias,[27] Javier Benitez,[28] Susan Neuhausen,[29] Hoda Anton-Culver,[30] Argyrios Ziogas,[30] Christina C. Dur,[31] Hermann Brenner,[32] Heiko Müller,[32] Volker Arndt,[32] Christa Stegmaier,[33] Alfons Meindl,[34] Rita K. Schmutzler,[35] Christoph Engel,[36] Nina Ditsch,[37] Hiltrud Brauch,[38] Thomas Brüning,[39] Yon-Dschun Ko,[40] The GENICA Network, Heli Nevanlinna,[41] Taru A. Muranen,[41] Kristiina Aittomäki,[42] Carl Blomqvist,[43] Keitaro Matsuo,[44] Hidemi Ito,[45] Hiroji Iwata,[46] Yasushi Yatabe,[47] Thilo Dörk,[48] Sonja Helbig,[48] Natalia V. Bogdanova,[49] Annika Lindblom,[50] Sara Margolin,[51] Arto Mannermaa,[52] Vesa Kataja,[53] Veli-Matti Kosma,[52] Jaana M. Hartikainen,[52] Georgia Chenevix-Trench,[3] kConFab Investigators, Australian Ovarian Cancer Study Group, Anna H. Wu,[54] Chiu-chen Tseng,[54] David Van Den Berg,[54] Daniel O. Stram,[54] Diether Lambrechts,[55,56] Bernard Thienpont,[55,56] Marie-Rose Christiaens,[57] Ann Smeets,[58] Jenny Chang-Claude,[59] Anja Rudolph,[59] Petra Seibold,[59] Dieter Flesch-Janys,[60] Paolo Radice,[61] Paolo Peterlongo,[62] Bernardo Bonanni,[63] Loris Bernard,[64] Fergus J. Couch,[65] Janet E. Olson,[66] Xianshu Wang,[65] Kristen Purrington,[67] Graham G. Giles,[5,68] Gianluca Severi,[5,68] Laura Baglietto,[68] Catriona McLean,[69] Christopher A. Haiman,[54] Brian E. Henderson,[54] Fredrick Schumacher,[54] Loic Le Marchand,[70] Jacques Simard,[71] Mark S. Goldberg,[72] France Labrèche,[73] Martine Dumont,[71] Soo-Hwang Teo,[74,75] Cheng-Har Yip,[75] Sze-Yee Phuah,[74,75] Vessela Kristensen,[76,77] Grethe Grenaker Alnæs,[76] Anne-Lise Børresen-Dale,[76,77] Wei Zheng,[78] Sandra Deming-Halverson,[78] Martha Shrubsole,[78] Jirong Long,[78] Robert Winqvist,[79] Katri Pylkäs,[79] Arja Jukkola-Vuorinen,[80] Saila Kauppila,[80] Irene L. Andrulis,[81,82] Julia A. Knight,[81,83] Gord Glendon,[81] Sandrine Tchatchou,[81] Peter Devilee,[84] Robert A.E.M. Tollenaar,[85] Caroline M. Seynaeve,[86] Montserrat García-Closas,[14,87] Jonine Figueroa,[88] Stephen J. Chanock,[88] Jolanta Lissowska,[89] Kamila Czene,[90] Hartef Darabi,[90] Kimael Eriksson,[90] Maartje J. Hooning,[86] John W.M. Martens,[86] Ans M.W. van den Ouweland,[91] Carolien H.M. van Deurzen,[92] Per Hall,[90] Jingmei Li,[93] Jianjun Liu,[93] Keith Humphreys,[90] Xiao-Ou Shu,[78] Wei Lu,[94] Yu-Tang Gao,[95] Hui Cai,[78] Angela Cox,[96] Malcolm W.R. Reed,[96] William Blot,[78,96] Lisa B. Signorello,[78,97] Qiuyin Cai,[78] Paul D.P. Pharoah,[2] Maya Ghoussaini,[2] Patricia Harrington,[2] Jonathan Tyrer,[2] Daehee Kang,[98,99] Ji-Yeob Choi,[99] Sue K. Park,[98] Dong-Young Noh,[100] Mikael Hartman,[101] Miao Hui,[101] Wei-Yen Lim,[101] Shaik A. Buhari,[102] Ute Hamann,[103] Asta Försti,[104] Thomas Rüdiger,[105] Hans-Ulrich Ulmer,[106] Anna Jakubowska,[107] Jan Lubinski,[107] Katarzyna Jaworska,[107,108] Katarzyna Durda,[107] Suleeporn Sangrajrang,[109] Valerie Gaborieau,[110] Paul Brennan,[110] James McKay,[110] Celine Vachon,[66] Susan Slager,[66] Florentia Fostira,[111] Robert Pilarski,[112] Chen-Yang Shen,[113,114] Chia-Ni Hsiung,[113] Pei-Ei Wu,[114] Ming-Feng Hou,[115] Anthony Swerdlow,[89,116] Alan Ashworth,[14,116] Nick Orr,[14,116] Minouk J. Schoemaker,[89] Bruce A.J. Ponder,[1] Alison M. Dunning,[2] and Douglas F. Easton[2]

[1]CRUK Cambridge Institute and Department of Oncology, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK; [2]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care and Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK; [3]Department of Genetics, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia; [4]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia; [5]Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, Melbourne, VIC 3010, Australia; [6]Department of Pathology, The University of Melbourne, Melbourne, VIC 3010, Australia; [7]Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, 1066 CX Amsterdam, the Netherlands; [8]Warwick Medical School, University of

Warwick, Coventry CV4 7AL, UK and Institute of Population Health, University of Manchester, Manchester M13 9PL, UK; [9]Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK; [10]Ministry of Public Health, Bangkok 10400, Thailand; [11]University Breast Center Franconia, Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, 91054 Erlangen, Germany; [12]David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, CA 90095, USA; [13]London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK; [14]Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, SW3 6JB, UK; [15]Division of Cancer Studies, NIHR Comprehensive Biomedical Research Centre, Guy's & St. Thomas' NHS Foundation Trust in partnership with King's College London, London SE1 9RT, UK; [16]Wellcome Trust Centre for Human Genetics and Oxford Biomedical Research Centre, University of Oxford, Oxford OX3 7BN, UK; [17]Clinical Science Institute, University Hospital Galway, Galway, Ireland; [18]Department of Obstetrics and Gynecology, University of Heidelberg, 69120 Heidelberg, Germany; [19]National Center for Tumor Diseases, University of Heidelberg, 69120 Heidelberg, Germany; [20]Molecular Epidemiology Group, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; [21]Inserm (National Institute of Health and Medical Research), CESP (Center for Research in Epidemiology and Population) and University Paris-Sud, UMRS 1018, 94807 Villejuif, France; [22]Université Paris Sorbonne Cité, UMR-S775 Inserm, 75270 Paris Cedex 06, France; [23]Copenhagen General Population Study and Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, University of Copenhagen, Copenhagen, 2730 Herlev, Denmark; [24]Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Copenhagen, 2730 Herlev, Denmark; [25]Genetic & Molecular Epidemiology Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain; [26]Servicio de Oncología Médica, Hospital Universitario La Paz, 28046 Madrid, Spain; [27]Servicio de Cirugía General y Especialidades, Hospital Monte Naranco, 33012 Oviedo, Spain; [28]Human Genetics Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain and Centro de Investigación en Red de Enfermedades Raras (CIBERER), 46010 Valencia, Spain; [29]Beckman Research Institute of City of Hope, Duarte, CA 91010, USA; [30]Department of Epidemiology, University of California, Irvine, Irvine, CA 92697, USA; [31]Cancer Prevention Institute of California, Fremont, CA 94538, USA; [32]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; [33]Saarland Cancer Registry, 66024 Saarbrücken, Germany; [34]Division of Gynaecology and Obstetrics, Technische Universität München, 81675 Munich, Germany; [35]Division of Molecular Gyneco-Oncology, Department of Gynaecology and Obstetrics at the University Cologne, 50931 Cologne, Germany; [36]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, 04107 Leipzig, Germany; [37]Department of Gynaecology and Obstetrics, Ludwig-Maximilians-Universität, 80539 Munich, Germany; [38]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, 70376 Stuttgart, Germany and University of Tübingen, 72074 Tübingen, Germany; [39]Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), 44789 Bochum, Germany; [40]Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, 53113 Bonn, Germany; [41]Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, 00029 HUS, Finland; [42]Department of Clinical Genetics, Helsinki University Central Hospital, Helsinki, 00029 HUS, Finland; [43]Department of Oncology, Helsinki University Central Hospital, Helsinki, 00029 HUS, Finland; [44]Department of Preventive Medicine, Kyushu University Faculty of Medical Sciences, Fukuoka 812-8582, Japan; [45]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya 464-8681, Japan; [46]Department of Breast Oncology, Aichi Cancer Center Hospital, Nagoya 464-8681, Japan; [47]Department of Pathology and Molecular Diagnostics, Aichi Cancer Center Hospital, Nagoya 464-8681, Japan; [48]Department of Obstetrics and Gynaecology, Hannover Medical School, 30625 Hannover, Germany; [49]Department of Radiation Oncology, Hannover Medical School, 30625 Hannover, Germany; [50]Department of Molecular Medicine and Surgery, Karolinska Institutet, 17177 Stockholm, Sweden; [51]Department of Oncology-Pathology, Karolinska Institutet, 17177 Stockholm, Sweden; [52]School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine and Cancer Center of Eastern Finland and University of Eastern Finland, and Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, 70211 Kuopio, Finland; [53]School of Medicine, Institute of Clinical Medicine, Oncology, University of Eastern Finland and Cancer Center, Kuopio University Hospital, 70211 Kuopio, Finland; [54]Department of Preventive Medicine, University of Southern California Keck School of Medicine, Los Angeles, CA 90089, USA; [55]Vesalius Research Center (VRC), VIB, 3000 Leuven, Belgium; [56]Department of Oncology, University of Leuven, 3000 Leuven, Belgium; [57]Oncology Department, University Hospital Gasthuisberg, 3000 Leuven, Belgium; [58]Universital Hospital Gasthuisberg, 3000 Leuven, Belgium; [59]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; [60]Department of Cancer Epidemiology/Clinical Cancer Registry and Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, 20246 Hamburg, Germany; [61]Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), 20133 Milan, Italy; [62]IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, 20139 Milan, Italy; [63]Division of Cancer Prevention and Genetics, Istituto Europeo di Oncologia (IEO), 20141 Milan, Italy; [64]Department of Experimental Oncology, Istituto Europeo di Oncologia (IEO), 20141 Milan, Italy and Cogentech Cancer Genetic Test Laboratory, 20139 Milan, Italy; [65]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, USA; [66]Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA; [67]Karmanos Cancer Center, Detroit, MI 48201, USA; [68]Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, VIC 3053, Australia; [69]Anatomical Pathology, The Alfred Hospital, Melbourne, VIC 3181, Australia; [70]Epidemiology Program, Cancer Research Center, University of Hawaii, Honolulu, HI 96813, USA; [71]Cancer Genomics Laboratory, Centre Hospitalier Universitaire de Québec and Laval University, Quebec City, QC G1V 4G2, Canada; [72]Department of Medicine, McGill University, Montreal, QC H3A 1A1, Canada and Division of Clinical Epidemiology, McGill University Health Centre, Royal Victoria Hospital, Montreal, QC H3A 1A1, Canada; [73]Département de médecine sociale et préventive, Département de santé environnementale et santé au travail, Université de Montréal, Montreal, QC H3A 3C2, Canada; [74]Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Subang Jaya, 47500 Selangor, Malaysia; [75]Breast Cancer Research Unit, University Malaya Cancer Research Institute, University Malaya Medical Centre, 50603 Kuala Lumpur, Malaysia; [76]Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Radiumhospitalet, 0310 Oslo, Norway; [77]Faculty of Medicine (Faculty Division Ahus), University of Oslo (UiO), 0318 Oslo, Norway; [78]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN 37203, USA; [79]Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry and Biocenter Oulu, University of Oulu, Oulu University Hospital, 90014 Oulu, Finland; [80]Department of Oncology, Oulu University Hospital, University of Oulu, 90014 Oulu, Finland; [81]Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada; [82]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada; [83]Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON M5S 1A8, Canada; [84]Department of Human Genetics & Department of Pathology, Leiden University Medical Center, 2300 RC Leiden, the Netherlands; [85]Department of Surgical Oncology, Leiden University Medical Center, 2300 RC Leiden, the Netherlands; [86]Department of Medical Oncology, Erasmus MC Cancer Institute, 3075 EA Rotterdam, the Netherlands; [87]Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK; [88]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD 20892, USA; [89]Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Memorial Cancer Center & Institute of Oncology, Warsaw 02-781, Poland; [90]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden; [91]Department of Clinical Genetics, Erasmus University Medical Center, 3008 AE Rotterdam, the Netherlands; [92]Department of Pathology, Erasmus University Medical Center, 3008 AE Rotterdam, the Netherlands; [93]Human Genetics Division, Genome Institute of Singapore, Singapore 138672, Singapore; [94]Shanghai Center for Disease Control and Prevention, Shanghai 200336, China; [95]Department of Epidemiology, Shanghai Cancer Institute, Shanghai 200032, China; [96]CRUK/YCR Sheffield Cancer Research Centre, Department of Oncology, University of Sheffield, Sheffield, S10 2RX, UK; [97]International Epidemiology Institute, Rockville, MD 20850, USA; [98]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul 110-799, Korea; [99]Department of Biomedical Sciences, Seoul National University Graduate School, Seoul 110-799, Korea; [100]Department of Surgery, Seoul National University College of Medicine, Seoul 110-799, Korea; [101]Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117597, Singapore and National University Health System, Singapore 119228, Singapore; [102]Division of General Surgery, National University Health System, Singapore 119228; [103]Molecular Genetics of Breast Cancer, German Cancer Research Centre (DKFZ), 69120 Heidelberg, Germany; [104]Center for Primary Health Care Research, Clinical Research Centre (CRC), Skanes universitetssjukhus, 20502 Malmö, Sweden; [105]Institute of Pathology, Städtisches Klinikum Karlsruhe, 76133 Karlsruhe,

The 10q26 locus in the second intron of *FGFR2* is the locus most strongly associated with estrogen-receptor-positive breast cancer in genome-wide association studies. We conducted fine-scale mapping in case-control studies genotyped with a custom chip (iCOGS), comprising 41 studies (n = 89,050) of European ancestry, 9 Asian ancestry studies (n = 13,983), and 2 African ancestry studies (n = 2,028) from the Breast Cancer Association Consortium. We identified three statistically independent risk signals within the locus. Within risk signals 1 and 3, genetic analysis identified five and two variants, respectively, highly correlated with the most strongly associated SNPs. By using a combination of genetic fine mapping, data on DNase hypersensitivity, and electrophoretic mobility shift assays to study protein-DNA binding, we identified rs35054928, rs2981578, and rs45631563 as putative functional SNPs. Chromatin immunoprecipitation showed that FOXA1 preferentially bound to the risk-associated allele (C) of rs2981578 and was able to recruit ERα to this site in an allele-specific manner, whereas E2F1 preferentially bound the risk variant of rs35054928. The risk alleles were preferentially found in open chromatin and bound by Ser5 phosphorylated RNA polymerase II, suggesting that the risk alleles are associated with changes in transcription. Chromatin conformation capture demonstrated that the risk region was able to interact with the promoter of *FGFR2*, the likely target gene of this risk region. A role for FOXA1 in mediating breast cancer susceptibility at this locus is consistent with the finding that the *FGFR2* risk locus primarily predisposes to estrogen-receptor-positive disease.

Introduction

Multiple genome-wide association studies (GWASs) have identified common variants on 10q26 associated with estrogen-receptor-positive (ER^+) breast cancer. These SNPs are the most strongly associated common variants identified for breast cancer[1–8] and map to the second intron of *FGFR2* (MIM 176943). Previous fine-scale mapping of this locus, together with analysis of evolutionary conservation and accessible chromatin, pointed to SNP rs2981578 being the most likely candidate causative variant.[8,9] Biochemical analysis of protein-DNA interactions at the risk locus also suggested rs2981578 as a functional variant, with the cancer-risk allele preferentially binding OCT1/RUNX2 in vitro.[10] siRNA experiments indicated that changes in RUNX2 can affect *FGFR2* expression levels.[11] However, in vivo chromatin immunoprecipitation (ChIP) assays found low levels of enrichment for OCT1/RUNX2 at this site,[10] suggesting that the mechanism by which risk is conferred at this locus has not yet been resolved. Experiments aimed at identifying the target gene(s) of this risk locus have implicated *FGFR2*, but the effects of the potential risk SNPs are still debated.[10,12,13]

Here we present the results of comprehensive fine-scale mapping of the *FGFR2* locus by using dense SNP genotyping in 52 case-control studies from populations of European, Asian, and African American ancestry within the Breast Cancer Association Consortium (BCAC). Furthermore, we examine allele-specific binding by FOXA1, ERα, E2F1, and RNA polymerase II to the candidate causal risk SNPs and propose a mechanism by which these SNPs may function to increase the risk of ER^+ disease.

Material and Methods

Genetic Mapping

Genotyping was conducted with a custom Illumina iSelect array (iCOGS) (for details see Michailidou et al.[6]). For this project, we identified SNPs across a 500 kb interval (positions 123,210,010–123,710,010 [NCBI build 37 assembly]) from the 1000 Genomes Project. This interval encompassed all known SNPs correlated ($r^2 > 0.1$) with the candidate causal variant, rs2981578. At the time of the chip design (March 2010), the 1000 Genomes Project had cataloged 3,431 variants with a minor allele frequency (MAF) > 2% in Europeans. From this catalog, we selected all SNPs correlated with rs2981578 ($r^2 > 0.1$) plus a set of SNPs tagging all remaining variants (at $r^2 > 0.9$). In total, 490 SNPs were designed for the iCOGS chip, of which 438 were successfully genotyped and passed quality control (see Michailidou et al.[6] for details).

After quality-control exclusions, genotypes were available for 89,050 individuals of European ancestry from 41 studies, 13,983 individuals from 9 Asian studies, and 2,048 individuals from 2 African ancestry studies.[6] All studies were approved by the relevant local ethics review committee and subjects gave informed consent.

Statistical Analysis

The genotype data were first used to estimate genotypes for other common variants across the region in the study subjects by imputation, with IMPUTE v.2.2 and the March 2012 release of the 1000 Genomes Project as reference panel. Genotypes at 2,291 SNPs could be imputed with imputation $r^2 > 0.3$. Per-allele ORs for each SNP were estimated by logistic regression, including study and principal components (seven in Europeans, two in Asians, and two in African Americans) as covariates, to allow for potential population stratification as previously described.[6] To determine the minimal number of SNPs independently associated with breast cancer for each ethnicity, forward stepwise logistic regression analysis was applied (with the R function *step*) to all SNPs with a MAF > 0.02 for which evidence of association

Germany; [106]Frauenklinik der Stadtklinik Baden-Baden, 76532 Baden-Baden, Germany; [107]Department of Genetics and Pathology, Pomeranian Medical University, u. Polabska 4, 70-115 Szczecin, Poland; [108]Postgraduate School of Molecular Medicine, Warsaw Medical University, ul. Żwirki i Wigury 61, 02-091 Warsaw, Poland; [109]National Cancer Institute, Bangkok 10400, Thailand; [110]International Agency for Research on Cancer, 69372 Lyon Cedex 08, France; [111]Molecular Diagnostics Laboratory, INRASTES, NSCR Demokritos, Athens 15310, Greece; [112]Department of Internal Medicine, James Comprehensive Cancer Center, Ohio State University, Columbus, OH 43210, USA; [113]Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan; [114]Colleague of Public Health, China Medical University, Taichung 40402, Taiwan; [115]Taiwan Biobank and Cancer Center and Department of Surgery, Kaohsiung Medical University Chung-Ho Memorial Hospital, Kaohsiung, Taiwan; [116]Division of Breast Cancer Research, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK
*Correspondence: kerstin.meyer@cruk.cam.ac.uk

(p value $< 1 \times 10^{-4}$) was observed in the single-SNP analysis. The p value for each SNP, after adjustment for all other SNPs, was determined by a Wald test. Haplotype-specific ORs and confidence limits were estimated with the haplo.stats package in R, with adjustment for study and principal components.

Cell Lines

Breast cancer cell lines MDA-MB-134, ZR-75-1, T47D, and MCF-7 were grown in RPMI medium with 10% FCS and antibiotics under standard conditions. These cell lines were from the CRUK Cambridge Institute's tissue culture collection. The normal breast epithelial cell line BRE80 (provided as a gift from Roger Reddel, CMRI, Sydney) was grown in DMEM/F12 medium with 5% horse serum (HS), 10 μg/ml insulin, 0.5 μg/ml hydrocortisone, 20 ng/ml epidermal growth factor, 100 ng/ml cholera toxin, and antibiotics.

Genotyping

Genomic DNA sequences from ZR-75-1, T47D, and MDA-MB-134 cells were amplified with primer pairs P1, P2, and P3. Resulting fragments were isolated and directly sequenced by GATC-Biotech. Additional cell lines were genotyped with fluorescent 5′ exonulease assay (TaqMan, predesigned assay) or by sequencing. Relevant genotypes are listed in Table S2 available online.

Electrophoretic Mobility Shift Assays

Small-scale nuclear extracts and bandshifts were carried out as previously described[10] and oligonucleotide sequences used in the assays are listed in Table S1. Competitor oligonucleotides were used at 10-, 30- and 100-fold molar excess as stated. Additional oligonucleotides used as competitors in Figure S2 were as listed on the Santa Cruz website.

Chromatin Immunoprecipitation

ChIP experiments were carried out as previously described.[14] DNA was quantitated with Quant-IT and equal amounts of precipitate and input used in RT-PCR reaction with SYBR green master mix on a 7900HT RT-PCR System (Applied Biosystems). Primers are given in Table S1. Allele-specific PCR was carried out with TaqMan Genotyping Assays (predesigned assays, ABI). Polyclonal antibodies against FOXA1 (ab5089 and ab23738, Abcam, UK at 1:1 ratio), ERα (sc543x, Santa Cruz Biotech), and E2F1 (C-20) (sc-193-x, Santa Cruz Biotech) were used in ChIP experiments. All values obtained are normalized to input and enrichment is given relative to the negative *CCND1* control.[15] To account for the slight variations in the levels of FOXA1 and E2F1 protein (Figure S1) and efficiency of the ChIP in the different cell lines, positive controls were included in each ChIP experiment: the *GREB1* (MIM 611736) promoter (pGREB)[16] for FOXA1, *MFAP1* (MIM 600215) for E2F1, and heme oxygenase (*HMOX1* [MIM 141250]) for RNA polymerase II. Antibodies for Ser5P RNA polymerase II (ab5131) were also obtained from Abcam, UK. In these experiments primer pairs from the genomic region of 8q24 (see Table S1) were used as negative control. Each ChIP has yielded similar results in at least two independent experiments. For the rs2981578 TaqMan assay, titrations were carried out to show that Ct values were directly proportional to input for each of the alleles. The error bars denote the standard deviation in three technical replicates. The data are also presented in the form of allelic discrimination (AD) plots. The AD plots are obtained from a TaqMan assay in which two different fluorophores are each linked to a probe detecting the two different alleles. Amplification of each allele was followed with an Applied Biosystems Real Time PCR machine (7900HT) and the data analyzed with the SDS software. The SDS software converts the raw data to fluorescence intensity for each allele and then plots the results as a scatter graph of allele X versus allele Y.

FAIRE

Formaldehyde-assisted isolation of DNA regulatory elements (FAIRE) relies on crosslinking of nucleosomes to DNA, with1% formaldehyde for 10 min, followed by the isolation of noncrosslinked DNA. The isolated DNA is enriched for regulatory elements that are in an open chromatin conformation.[17] FAIRE was carried out in MDA-MB-134 cells grown in full medium. Sonicated genomic T47D and ZR-75-1 DNA was included in the genotyping assay to indicate the position of risk and nonrisk homozygotes in the allelic discrimination plots.

Chromatin Conformation Capture

Chromatin conformation capture (3C) libraries were generated with EcoRI as described previously.[18] 3C interactions were quantitated by real-time PCR (qPCR) with primers designed within EcoRI restriction fragments (Table S3). qPCR was performed on a RotorGene 6000 with MyTaq HS DNA polymerase (Bioline) with the addition of 5 mM of Syto9, annealing temperature of 66°C, and extension of 30 s. 3C analyses were performed in two independent experiments with each experiment quantified in triplicate. BAC clone RP11-62L18 covering the 10q26 region was used to create an artificial library of ligation products in order to normalize for PCR efficiency. Data were normalized to the signal from the BAC clone library and, between cell lines, by reference to a region at within *GAPDH* (MIM 138400). All qPCR products were electrophoresed on 2% agarose gels, gel purified, and sequenced to verify the 3C product.

siRNA Transfections

Cells were grown to 50% confluence in 6-well plates. Transfections were carried out according to manufacturer's instructions with 10 μl lipofectamine RNAiMAX (Invitrogen) in a total of 3 ml OptiMEM medium (Invitrogen), with siRNAs at a final concentration of 50 nM. A custom siRNA against *FOXA1* (5′-GAGAGAA AAAAUCAACAGC-3′)[16,19] and the On-TARGET plus nontargeting pool (D001810) negative control were obtained from Thermo Fisher Dharmacon. After 6–24 hr the transfection mix was replaced with normal growth medium and cells were harvested after 48 hr by scraping into cold PBS. After washing in PBS at 4°C, samples were split in two. For protein analysis, samples were resuspended in RIPA buffer plus protease inhibitors (Roche) and frozen on dry ice and depletion of the relevant protein was confirmed by immunoblot. For RNA isolation, samples were resuspended in 1× RLT and the Allprep kit (QIAGEN) was used to purify nucleic acids.

Gene Expression Analysis

Gene expression in siRNA-transfected samples was analyzed by RT-PCR for *FOXA1* (MIM 602294) and normalized against the housekeeping gene *DGUOK* (MIM 601465) with SYBR green PCR Mastermix. Oligonucleotides were designed with Primer 3 (v.0.4.0) and sequences are given in Table S1. *FGFR2* mRNA levels were measured by TaqMan assay (*FGFR2* Taqman assay Hs00240796m1) and normalized against *GAPDH* (Taqman predeveloped assay 4333764T) from Applied Biosystems. Expression of *FOXA1* and *FGFR2* after transfection of siRNA is given relative to

the control transfected cells. All transfections were carried out in triplicate.

Gene Expression Correlations and Statistical Analysis

Evidence for association between gene expression levels and SNP genotypes was evaluated with ANOVA. Gene expression data was obtained from the METABRIC sample collection.[20] Genotypes for rs2981578 and rs35054928 were determined with the probes A-8465362 and A-8444843 on the Illumina SNP 6 array.

Results

Fine-Scale Genetic Mapping

In total, 438 genotyped and 2,291 well-imputed SNPs were considered, of which 392 SNPs were associated with breast-cancer risk in at least one ethnic group at $p < 10^{-4}$ (Table S4 and Figure 1). All associations were confined to ER^+ disease, with no evidence of association with ER-negative (ER^-) breast cancer (Table S5). Gene annotations and chromatin modifications across the genotyped region are shown in Figure 1D.

Results from the European Ancestry Studies

A total of 375 SNPs were associated with breast-cancer risk in European woman at $p < 10^{-4}$ (Table S4). Stepwise logistic regression identified three independent association signals. Within each signal we identified most likely causal variants after excluding all variants having a likelihood ratio < 1/100 relative to the best-associated SNP in each signal (shown in Table 1 for overall breast cancer and Table 2 for ER^+ disease).

Within signal 1 the strongest association was with the insertion/deletion rs35054928 (MAF = 0.44; OR per C [insertion] allele = 1.27; 95% CI 1.24–1.29, p value 6.8×10^{-131}, p value 0.06 after adjustment for SNPs in other signals for all tumors, 0.009 for ER^+ tumors), but a further four SNPs were potentially causal with a likelihood ratio greater than 1/100. Signal 2 contains a single variant rs45631563 (MAF = 0.04; OR per A [minor] allele = 0.80; 95% CI 0.76–0.85, p value 3.8×10^{-15}, adjusted p values: 3.09×10^{-11} for all tumors, 1.26×10^{-11} for ER^+ tumors). Signal 3 contained two correlated variants ($r^2 = 0.99$): rs2981578 (MAF = 0.50; OR per C allele = 1.24; 95% CI 1.21–1.26; p value: 1.6×10^{-106}, conditional p value: 3.97×10^{-8} for all tumors, 1.98×10^{-6} for ER^+) and insertion/deletion variant rs45631539.

Genotypes of potentially causal SNPs in signals 1 and 3 were correlated (e.g., $r^2 = 0.79$ for rs35054928 and rs2981578) but the remaining signal 2 SNP was not correlated with the others (see Table S6).

Results from the Asian Ancestry Studies

A total of 30 SNPs displayed association with breast-cancer risk in Asian woman at $p < 10^{-4}$ (Table S7). The top SNP in Asians was rs2912781, which is highly correlated with rs2981578 ($r^2 = 0.97$ in Asians and $r^2 = 0.98$ in Europeans), the top SNP in signal 3 in Europeans. Based on stepwise logistic regression analysis (Tables 1 and 2), only rs2981578 was independently associated with risk (conditional p value: 0.0001 in Asians). Although SNP rs2912781 was also strongly associated with breast cancer in Europeans, it was excluded as a likely causative variant because it had a likelihood ratio < 1/500 relative to SNP rs2981578, the top signal 3 SNP in Europeans ($r^2 = 0.97$ in Asians and $r^2 = 0.98$ in Europeans).

Results from the African Ancestry Studies

One SNP displayed association with breast cancer in African woman at $p < 10^{-4}$ (Table S4). This SNP, rs74992784 (MAF = 0.02; OR ER^+ per minor allele = 2.05; 95% CI 1.41–2.98; p value 0.0001), maps approximately 200 kb upstream of *FGFR2*, in an intron of *ATE1* (MIM 607103), is not polymorphic in Europeans or Asians, and is not in LD with any of the other risk SNPs. The second strongest association was with SNP rs11200017 (MAF = 0.05; OR ER^+ per minor allele = 0.62; 95% CI 0.45–0.85; p value 0.003). In Europeans this SNP was associated with increased risk (MAF = 0.2; OR ER^+ per minor allele = 1.16; 95% CI 1.13–1.19; p value 4.9×10^{-34}). However, it is weakly correlated with the top SNPs in signals 1 and 3 and not significant after adjustment for these SNPs.

Haplotype Analysis

Haplotype analysis was performed with one tag SNP from each of the three independent risk signals detected in Europeans for ER^+ breast cancer (Table 2). In Europeans we observed four haplotypes with frequencies >1%, each associated with a different level of risk (Table 3). The highest risk was conferred by the haplotype (haplotype 3) carrying the risk alleles for both rs35054928 (signal 1) and rs2981578 (signal 3) (OR: 1.33; 95% CI 1.31–1.34, compared to the baseline haplotype, carrying the common allele at all three SNPs). Although these two SNPs mark two independent risk signals, they are physically very close (only 120 bp apart). Haplotype 1, which carries the risk allele at rs2981578 but not rs35054928, was associated with an intermediate increased risk (OR 1.14; 95% CI 1.1–1.18). Haplotype 2 carries the rare allele for rs45631563 (signal 2) in addition to the risk allele at rs2981578 and was associated with a reduced risk (OR 0.86: 95% CI 0.83–0.88) relative to the baseline haplotype. rs45631563 lies approximately 5.6 kb from the other two risk SNPs.

Haplotype 3 also conferred the highest risk in Asians (OR 1.27: 95% CI 1.23–1.31). Haplotype 1 conferred an increased risk, but the OR was higher than in Europeans and was not significantly different from that conferred by haplotype 3 (OR 1.24; 95% CI 1.17–1.31). Haplotype 2 was not associated with a different risk, relative to the baseline haplotype, but it was relatively rare (frequency 1%) and the OR was consistent with that observed for Europeans (OR 0.96; 95% CI 0.81–1.14).

Given the difference in the results obtained for Asians and Europeans, we extended the haplotype analysis to

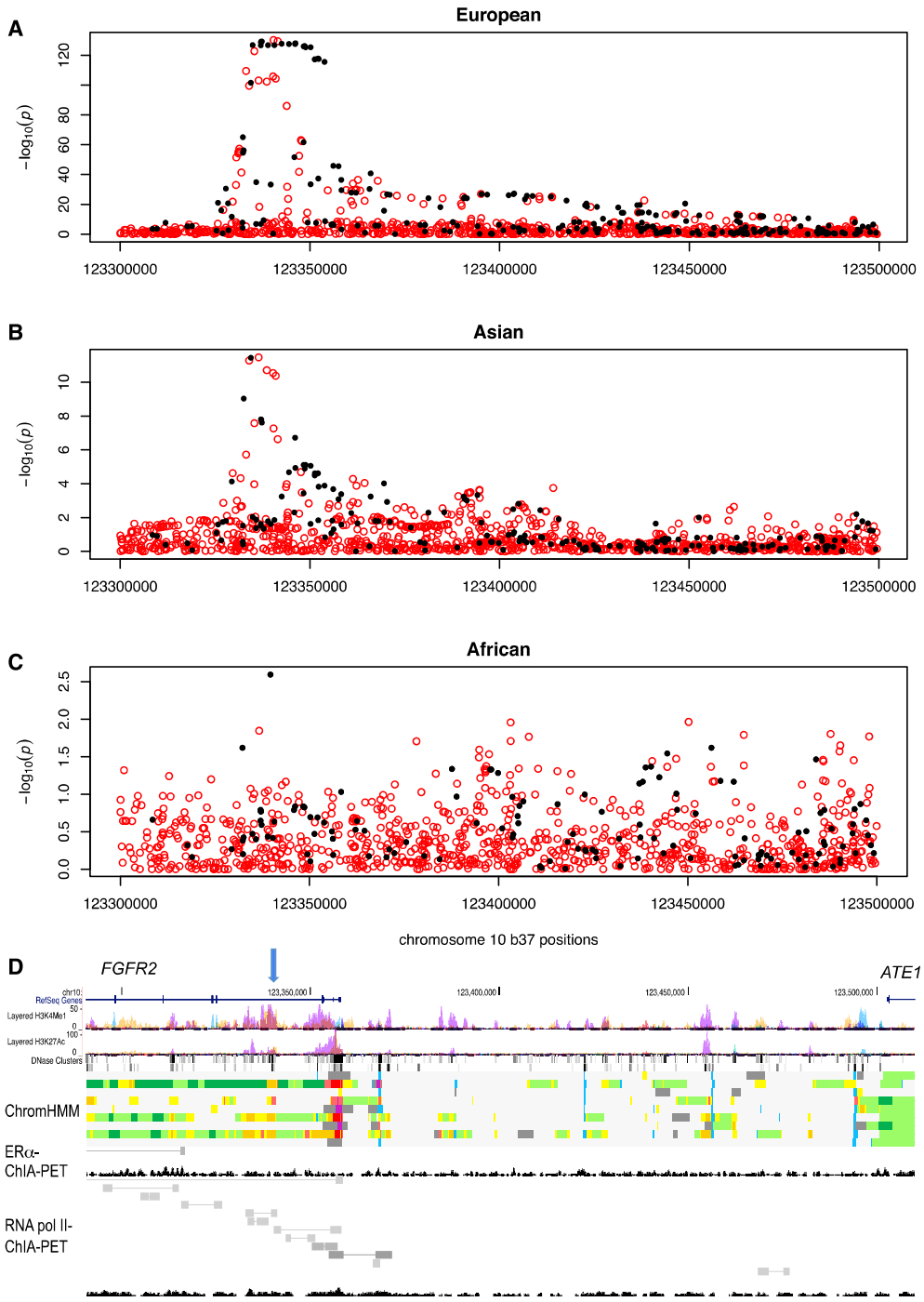

Figure 1. Manhattan Plot of the *FGFR2* Breast Cancer Risk Locus

(A–C) Genotyped and imputed SNPs are plotted based on their chromosomal position on the x axis and their overall p value (-log10 values) from (A) European BCAC studies, (B) Asian ancestry, and (C) African ancestry on the y axis.

(D) Chromatin configuration across the genotyped region from ENCODE. From top to bottom, lanes show RefSeq genes (*FGFR2* and *ATE1*), H3K4Me1 and H3K27Me1 histone modifications, DNase clusters, transcription factor ChIP, ChromHMM chromatin states in GM12878, H1-hESC, K562, HepG2, HUVEC, HMEC, HSMM, NHEK, and NHLF cells, ER-α ChIA-PET, and RNA polymerase II ChIA-PET in MCF-7 cells. The blue arrow highlights the position of the top SNP (rs35054928). ChromHMM color coding is as follows: bright red, active promoter; light red, weak promoter; purple, inactive/poised promoter; orange, strong enhancer; yellow, weak enhancer; blue, insulator; dark green, transcriptional transition; gray, repressed/heterochromatin.

include the top SNP from signal 1 in the African American analysis (Table S8). This revealed that the minor allele of rs11200017 was protective in all three ethnicities. It is more common in Asians than Europeans and is always found in combination with the minor allele of rs35054928. It might therefore partly explain why

Table 1. Candidate Functional Variants within the Three Independent Association Signals in Europeans and Asians: Association Results for Overall Breast Cancer

				Europeans		Asians			
Signal	SNP (Position)	Alleles	MAF	OR (95% CI)	p Trend (adjusted)[a]	OR (95% CI)	p Trend (adjusted)[a]	DHS[b] (MCF7)	Known TF Binding Site[b]
1	rs35054928 (123340431)	-/C	0.44	1.27 (1.24–1.29)	6.8×10^{-131} (0.06)	1.15 (1.10–1.21)	5.4×10^{-8} (0.771)	yes	E2F1/FOXA1
	rs34032268 (123341525)	C/A	0.42	1.27 (1.25–1.30)	3.5×10^{-130}	1.15 (1.09–1.21)	2.3×10^{-7}	no	no
	rs2981579 (123337335)	G/A	0.43	1.27 (1.24–1.29)	5.7×10^{-130}	1.16 (1.1–1.22)	2.4×10^{-8}	yes	no
	rs2912779 (123337182)	C/T	0.43	1.27 (1.24–1.29)	8.3×10^{-130}	1.16 (1.1–1.22)	1.7×10^{-8}	yes	no
	rs2912780 (123337117)	T/C	0.42	1.27 (1.24–1.29)	1.7×10^{-130}	1.16 (1.1–1.22)	1.6×10^{-8}	no	no
2	rs45631563 (123349324)	T/A	0.035	0.80 (0.76–0.85)	3.8×10^{-15} (3.1×10^{-11})	1.02 (0.75–1.39)	0.89 (0.450)	yes	no
3	rs2981578 (123340311)	T/C	0.497	1.24 (1.21–1.26)	1.6×10^{-106} (4.0×10^{-8})	1.19 (1.13–1.25)	2.9×10^{-11} (0.0001)	yes	E2F1/FOXA1
	rs45631539 (123341009)	-/ins	0.499	1.24 (1.21–1.26)	4.5×10^{-105}	1.19 (1.13–1.25)	4.2×10^{-11}	no	no

Abbreviations are as follows: MAF, minor allele frequency; DHS, DNase I hypersensitive site; TF, transcription factor; ins, insertion of TGGGAGGCCAAGG.
[a]After adjustment for the top SNPs in the other signals, in a logistic regression model including rs35054928, rs45631563, and rs2981578.
[b]Taken from the UCSC Genome Browser, hg19.

rs35054928 is not independently associated with risk in Asians, after adjustment for rs2981578.

Analysis of Protein-DNA Interactions

By using data from ENCODE, we examined DNase hypersensitivity and known transcription factor binding at each of the potential causative SNPs in risk signals 1, 2, and 3 (Table 1) and selected for further analysis five SNPs (rs35054928, rs2981579, rs2912779, rs2981578, and rs45631563) that were found to be present in a DNase hypersensitive site (DHS) in MCF-7 cells (UCSC Genome Browser), because confirmed GWAS hits have been shown to be enriched in open chromatin.[21] Electrophoretic mobility shift assays (EMSAs) were carried out for these variants. For rs35054928 we detected binding by two protein complexes. Competition experiments suggested that the strong (lower) band is likely to be E2F1 (Figure 2A), which has also been reported to bind at this site in vivo (UCSC Genome Browser), whereas the weaker band could be competed with SP1 oligonucleotides (Figure 2A). We observed protein interaction at rs2981579, but no difference between alleles. Competition experiments indicated that this oligonucleotide is bound by ERα in vitro (Figures S2A and S1B), but ChIP-seq data suggest that this site is not occupied by ERα in vivo (Figure S3).[22] At rs2912779, EMSA again showed no allele-specific differences in binding by an as yet unidentified protein (Figure S2A). For rs2981578 in risk signal 3, we previously reported allele-specific binding of OCT1/RUNX2 in an ER^- cell line. However, after further analysis in ER^+ cell lines, we found that, in addition to the OCT1/RUNX2 bands, there was strong binding by FOXA1 in MCF-7 (Figure 3B) and also in T47D, ZR-75-1 (Figure S2C), and PMC42 nuclear extracts.[10] At rs45631563 (signal 2), we detected binding by a nuclear protein with a stronger affinity for the A allele (Figure 2C), but extensive competition analysis did not reveal the identity of this protein (Figures S2D and S2E).

Protein binding to rs11200017, the potential risk SNP in African Americans, was assessed bioinformatically. These analyses suggested that the G allele of this SNP preferentially forms GATA and ETS motifs, whereas the A allele generates a sequence homologous to winged-helix recognition motifs.

To examine transcription factor binding in vivo by ChIP, we genotyped ER^+ cell lines for the risk SNPs rs35054928 and rs2981578. The SNPs are correlated ($r^2 = 0.79$ in Europeans) and all cell lines examined were concordant for the risk alleles at these SNPs (see Table S3 for full genotypes). The following lines were chosen for further analysis: ZR-75-1 (rs2981578 C/C) (risk homozygote), MDA-MB-134 (C/T) (heterozygote), and T47D (T/T) (nonrisk homozygote). In all further allele-specific ChIP experiments, the TaqMan probe for rs2981578 was utilized as read-out for both SNPs, because the two SNPs are in closer proximity (120 bp) than the resolution limit of the ChIP assay (about 400 bp).

In FOXA1-ChIP assays, the C allele of rs2981578 present in ZR-75-1 (C/C) and MDA-MB-134 (C/T) was strongly enriched compared to a negative control from the *CCND1* locus (Figure 3A). Binding at this site was found to be even stronger than that observed for the positive control pGREB. However, the T allele of this site was only marginally enriched and ChIP in the heterozygous cell line MDA-MB-134 showed 8.5-fold greater enrichment

Table 2. Candidate Functional Variants within the Three Independent Association Signals in Europeans and Asians: Association Results for ER$^+$ Breast Cancer

		Europeans ER$^+$		Asians ER$^+$	
Signal	**SNP (Position)**	**OR (95% CI)**	**p Trend (adjusted)[a]**	**OR (95% CI)**	**p Trend (adjusted)[a]**
1	rs35054928 (123340431)	1.33 (1.30–1.36)	1.2×10^{-134} (0.009)	1.23 (1.16–1.31)	7.7×10^{-11} (0.81)
	rs34032268 (123341525)	1.33 (1.30–1.36)	3.1×10^{-132}	1.23 (1.15–1.31)	1.9×10^{-10}
	rs2981579 (123337335)	1.33 (1.30–1.36)	6.6×10^{-133}	1.23 (1.16–1.31)	5.8×10^{-11}
	rs2912779 (123337182)	1.33 (1.30–1.36)	7.3×10^{-133}	1.23 (1.15–1.31)	4.1×10^{-11}
	rs2912780 (123337117)	1.33 (1.30–1.36)	1.2×10^{-132}	1.23 (1.16–1.31)	4.2×10^{-11}
2	rs45631563 (123349324)	0.74 (0.69–0.79)	4.5×10^{-19} (1.26×10^{-11})	1.10 (0.73–1.66)	0.63 (0.22)
3	rs2981578 (123340311)	1.28 (1.25–1.31)	3.3×10^{-104} (1.98×10^{-6})	1.26 (1.19–1.34)	1.2×10^{-13} (0.0001)
	rs45631539 (123341009)	1.28 (1.25–1.31)	7.8×10^{-103}	1.26 (1.19–1.35)	1.8×10^{-13}

[a]After adjustment for the top SNPs in the other signals, in a logistic regression model including rs35054928, rs45631563, and rs2981578.

for the C allele. Moreover, allelic discrimination plots (Figure 3B) confirmed that FOXA1-precipitated DNA from this cell line clustered with the homozygous (C/C) ZR-75-1 samples, further demonstrating the strong allele-specific binding of FOXA1 at this SNP. Inspection of the position weight matrix for FOXA1 confirmed that rs2981578 overlaps a region of homology to the consensus binding site for this transcription factor (Figure 3F), especially in those base pairs contacting the major groove (RTTTR). Of interest, the allelic discrimination displayed by FOXA1 when binding to chromatin in vivo was not detectable in EMSAs via naked DNA (Figure 2B). We also examined allele-specific binding of RUNX2 in MDA-MB-134 cells but obtained either no or low enrichment with no evidence of allele-specific binding (data not shown).

ChIP-seq data in ZR-75-1 cells suggested the presence of an ERα binding site adjacent to the FOXA1 binding site[22] (Figure S3A). We hypothesized that the pioneer factor FOXA1 might be responsible for recruiting ERα to this site. We therefore carried out ER-ChIP in the same set of cell lines and found that ERα binding recapitulates FOXA1 binding. In both ZR-75-1 (C/C) and MDA-MB-134 (C/T), there was clear enrichment for the C allele (Figure 3C). However, total levels of binding were low, suggesting that the strong peak seen in ChIP-seq (Figure S3A) is generated by a distinct site (see below). The T allele showed no enrichment (Figure 3C), whereas the positive ChIP control (pGREB) displayed good enrichment in all cell lines tested (Figure 3D). The allelic discrimination plot (Figure 3E) again confirms the strong bias of binding in favor of the risk allele.

In order to exclude any effect of additional SNPs that might influence ER binding, we sequenced the region overlapping the FOXA1 and ERα ChIP-seq peaks. When comparing ZR-75-1, MDA-MB-134, and T47D, we found two additional SNPs that were polymorphic in these cell lines: rs111729099 (deletion of C) and rs11599804 (G/A). By using an EMSA with the ER consensus binding site, we examined whether sequences overlapping these SNPs could compete for ER binding in an allele-specific manner. Figure S3B shows that sequences overlapping rs111729099 did not compete for binding. In contrast, an oligonucleotide overlapping rs11599804 was able to compete for ERα binding, but both alleles behaved identically. We therefore conclude that the difference in ERα binding we observe at rs2981578 is probably due to differential binding of FOXA1 to this SNP, which is able to recruit ERα.

We next examined E2F1 binding at rs35054928 in the same cell lines. ChIP experiments with a nonallelic probe showed that there is a 2.4-fold enrichment of E2F1 binding at the risk allele (ZR-75-1), whereas little enrichment was seen at the nonrisk allele (T47D) (Figure 4A). In the heterozygous cell line MDA-MB-134, enrichment was low but nevertheless there was preferential binding of the risk allele in this cell line as shown in the allelic discrimination plot (Figures 4A and 4B). Figure 4C indicates weak homology of the sequence surrounding rs35054928 to the E2F1 consensus binding site. ChIP was also carried out with SP1 antibodies but no enrichment or allele-specific binding was detectable at this site (data not shown).

Having detected allele-specific binding by FOXA1 and E2F1, we examined the transcriptional activity of the locus by assaying binding of RNA polymerase II at the risk SNPs via ChIP. We used antibodies against the serine-5-phosphorylated (Ser5P) form of this enzyme, which catalyzes transcriptional initiation and elongation and has been shown to be a good read-out for transcriptional activity.[23] Figure 5 shows that Ser5P-Pol II can be precipitated at the rs2981578 site. In the heterozygous cell line MDA-MB-134, there was a 3.8-fold greater enrichment of the C over the T allele (Figure 5A), which is visualized in the allelic discrimination plot (Figure 5B). This finding suggests that the risk allele (C) increases transcription. The increased binding of FOXA1 and RNA polymerase II is also reflected in increased chromatin accessibility of the risk allele, as shown by allele-specific sequence retrieval in a FAIRE assay (Figure 5C).

Table 3. *FGFR2* Risk Haplotype Analysis in Europeans and Asians

	rs35054928[a]	rs2981578[b]	rs45631563[c]	Frequency	OR	p1 df	95% CIs
Haplotypes in Europeans ER$^+$							
1 (2.4)	1	2	1	0.025	1.14	3.6×10^{-04}	1.10–1.18
2 (2.6)	1	2	2	0.035	0.86	3.3×10^{-06}	0.83–0.88
3 (2.9)	2	2	1	0.428	1.33	1.7×10^{-125}	1.31–1.34
Rare	*	*	*	0.004			
Baseline	1	1	1	0.509			
Haplotypes in Asians ER$^+$							
1 (2.4)	1	2	1	0.090	1.24	1.1×10^{-04}	1.17–1.31
2 (2.6)	1	2	2	0.010	0.96	8.2×10^{-01}	0.81–1.14
3 (2.10)	2	2	1	0.453	1.27	4.1×10^{-13}	1.23–1.31
Rare	*	*	*	0.000			
Baseline	1	1	1	0.446			

In the first column, 1 indicates major genotype in Europeans and 2 indicates minor genotype in Europeans.
[a]1 indicates - allele, 2 indicates C allele.
[b]1 indicates T allele, 2 indicates C allele.
[c]1 indicates A allele, 2 indicates T allele.

si*FOXA1* Can Reduce *FGFR2* Expression

Previous studies have implicated *FGFR2* as a target of the 10q26 risk locus. To confirm a role of FOXA1 in *FGFR2* expression, we transfected T47D and ZR-75-1 cells with siRNA against *FOXA1* and examined *FOXA1* and *FGFR2* expression 48 hr after transfection. Figure 6 shows that *FGFR2* expression was repressed to 34% of its pretransfection level in T47D (Figure 6A) cells and to 12% in ZR-75-1 cells (Figure 6B). These results are consistent with FOXA1 having a more pronounced effect in cells carrying the risk genotype. However, *FGFR2* contains multiple additional FOXA1 binding sites that may contribute to the observed results. Transfection of si*E2F1* had little effect on *FGFR2* expression (data not shown), but it also caused a small upregulation of *FOXA1*, making these data hard to interpret.

Genotype-Expression Correlations

To examine the potential target gene(s) underlying the SNP associations, we used microarray data from 1,920 breast tumors (METABRIC[20]) to assess associations between the expression of genes within a 1Mb interval of the risk SNPs (including *FGFR2*, *ATE1* [MIM 607103], *NSMCE4A* [MIM 612987], and *TACC2* [MIM 605302]) and the presence of the risk genotypes at rs2981578 and rs35054928 by using the probes A-8465362 and A-8444843, respectively, of the Illumina SNP6 array (correlated to the two risk SNPs with $r^2 = 0.935$ in each case). No significant associations with expression were observed. Expression levels of *FGFR2* and *NSMCE4A* by genotype in both normal and tumor samples are shown in Figure S4. However, RNA Pol II ChIA-PET experiments in MCF-7 cells (Figure 1D) indicate that the putative regulatory region encompassing the top SNPs interacts with the *FGFR2* promoter, whereas no interactions with neighboring genes were detected, suggesting that *FGFR2* is the likely target.

To examine whether the risk SNPs might function in a restricted cell population, we then examined the distribution of FOXA1 protein in normal mammary epithelial tissue. Consistent with previous studies,[24] Figure S5 shows that in the human mammary gland, FOXA1 is found in the nucleus of a subset of luminal epithelial cells only. Given this restricted tissue distribution, in the future it may be possible to carry out more powerful genotype-expression association analyses by using microdissected epithelial tissue.

The Risk Region Interacts with the *FGFR2* Promoter

To provide further evidence for *FGFR2* being the target gene, we examined the physical interaction between the risk SNPs and the *FGFR2* promoter via a 3C assay. Figure 7 shows that the putative regulatory element encompassing the putative functional SNPs is able to interact with the promoter sequences of *FGFR2* in both ER$^+$ (MCF-7 and T47D) and ER$^-$ (BRE-80) cell lines, strongly supporting our conclusion that FGFR2 is the likely target gene of the risk locus.

Discussion

GWASs have now identified more than 70 breast-cancer risk loci.[6] However, our understanding of the mechanisms by which these loci confer risk is still limited, and for the large majority of GWAS hits neither the regulated target genes nor the causative SNPs are known. At the 10q26 risk locus, the risk region maps to the second intron of *FGFR2*, suggesting that *FGFR2* itself is the likely target

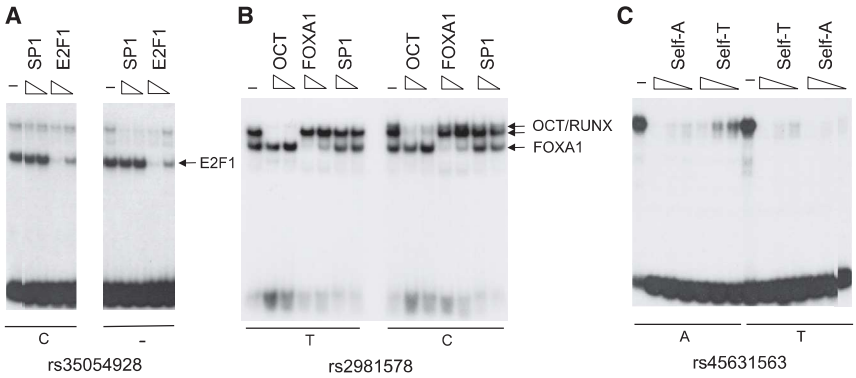

Figure 2. Electrophoretic Mobility Shift Assays for Candidate Causitive Variants
(A) rs35054928, (B) rs2981578, and (C) rs45631563 were assayed with MCF-7 nuclear extracts. Labels above each lane indicate inclusion of competitor oligonucleotides at 10- and 30-fold molar excess (A and B) and also at 100-fold molar excess (C). For each SNP, the common allele is shown on the left, the minor allele on the right.

gene, a hypothesis supported by previous studies reporting association between presence of the risk allele and *FGFR2* expression.[10,12] However, functional and biochemical studies of the likely causative SNPs have not fully explained the behavior of this risk locus.

Here we present extensive genetic fine mapping of this locus with >2,200 imputed and genotyped SNPs in a very large sample of breast-cancer cases and controls from the BCAC. We have identified three independent risk signals within this region, indicating that at least three variants are likely to be causally implicated. We find evidence of allele-specific transcription factor binding for the most strongly associated SNP in each of the three risk signals. Specifically, we detect in vivo allele-specific binding of E2F1 at rs35054928 and of FOXA1 and ERα at rs2981578 in a cell line heterozygous for these SNPs. TF binding is also reflected in allele-specific chromatin accessibility and recruitment of RNA polymerase II to this sites, which maps to a putative enhancer region. These results suggest that these two SNPs are most likely to be causally related to breast-cancer risk. At rs45631563, allele-specific binding by a nuclear protein was detected in vitro. We do not exclude the possibility that additional risk SNPs within each signal may also contribute to function.

We found some differences in the pattern of association between European and Asian women. Although clear evidence of independent associations was observed in Europeans for both rs35054928 and rs2981578, the analysis in Asians found no evidence for rs35054928, after adjustment for rs2981578. Haplotype analyses suggest that this difference might be due to the confounding effect of the third risk SNP (signal 2, rs45631563), for which the rare allele is protective in Europeans with a similar trend in Asians, where it is much rarer. The differences between Europeans and Asians might also reflect other modifying variants in the region, or other factors such as genetic background or differences in the distribution of tumor subtypes. We found no clear evidence of association for these SNPs in African Americans; however, the sample size was much smaller and therefore there was less statistical power in the fine-mapping analyses, particularly because a higher proportion of breast cancer cases in African American women are ER^-.

The pattern we report here, in which multiple independent variants in a region contribute to disease risk, has also been observed in fine-mapping studies of the *CCND1*[18] and *TERT*[25] breast-cancer susceptibility regions, suggesting that this may be a common feature. Because, as here, neighboring SNPs are often correlated, it is difficult to disentangle their independent effects, even in very large epidemiological studies and in the presence of highly statistically significant associations. The haplotype carrying the risk variants rs2981578 and rs35054928 confers an estimated relative risk of 1.33 (relative to the baseline haplotype), which puts it among the highest for a cancer susceptibility locus identified through GWASs. This explains why the *FGFR2* locus is the most readily identified risk locus in breast-cancer GWASs. One might speculate that, conversely, strong GWAS hits may tend to be the result of multiple causal variants.

Recent genome-wide analysis suggests that FOXA1 binding sites are enriched among breast-cancer susceptibility SNPs.[26,27] Our observation of allele-specific binding by FOXA1 to rs2981578 fits well with this result, but we demonstrate that additional factors contribute to cancer susceptibility. FOXA1 is known to act as a pioneer factor, able to open closed chromatin, thereby allowing the recruitment of additional factors, especially ERα.[28] Our observation of allele-specific binding by ERα at rs2981578 is fully consistent with such a model. It is tempting to speculate that FOXA1 may provide a similar function for the closely linked SNP rs35054928, for which we have shown allele-specific binding by E2F1. The fact that the FOXA1 binding allele is also present on a protective haplotype could also be explained in terms of its function as a pioneer factor. Once it has opened the chromatin, either activation or repressing factors might be recruited to its vicinity, as observed in gene expression studies after si*FOXA1* transfection.[16]

FOXA1 and ERα are part of a network of transcriptional master regulators conferring estrogen responsiveness.[16,24,29] Our findings that these two transcription factors are involved in allele-specific regulation of *FGFR2* are therefore fully consistent with our genetic association analyses that show a much stronger effect of this risk locus for ER^+ disease, with little or no association for ER^- disease.

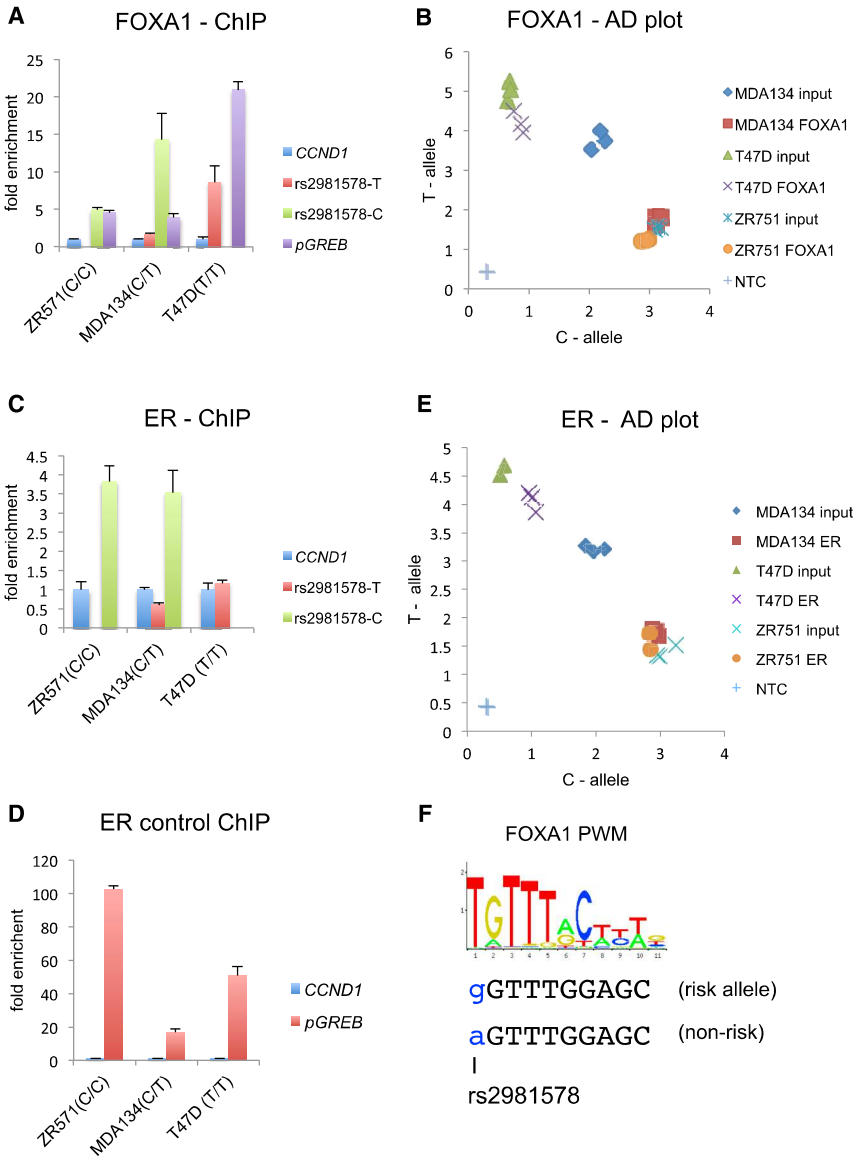

Figure 3. ChIP Assay at the rs2981578 Site

(A–E) FOXA1 (A) and ERα (C and D) ChIP-qPCR assays were carried out in the three cell lines ZR-75-1, MDA-MB-134, and T47D, each carrying a different genotype for this SNP. A primer pair from the *CCND1* locus served as negative control, and pGREB was used as positive control. Error bars show standard deviation for three technical replicates of a representative experiment. An allelic discrimination plot of the same experiments is shown for FOXA1 (B) and ERα (E) ChIPs. Abbreviation: NTC, nontemplate control.

(F) Position weight matrix, PWM, of FOXA1 from JASPAR, with homology to the (−) strand of risk (G) and nonrisk (A) alleles of rs2981578 shown in blue underneath.

E2F1 is a key transcription factor in the control of proliferation, in differentiation, and in the control of apoptosis. In breast-cancer samples its expression correlates well with other proliferation markers, is indicative of poor outcome,[30] and is independent of ER status. E2F1 may therefore in principle contribute to risk for developing both ER^+ and ER^- disease, but our association data suggest that the effect of rs35054928 on ER^- disease risk is small. In addition, we have previously shown that RUNX2 displays allele-specific binding at rs2981578, at least in vitro,[10] and may also contribute to risk in certain breast-cancer subtypes or at specific stages of development. High levels of RUNX2 have been found in triple-negative tumors,[31] but the association analyses suggest that the effect of rs2981578 on the development of ER^- tumors is also small.

A number of studies have addressed questions regarding the biological effect of the presence of the *FGFR2* risk SNPs: do the risk SNPs affect transcription and what are their likely target gene(s)? Our previous work has suggested that the risk allele of rs2981578 is able to drive transcription more strongly than the nonrisk allele.[10] This conclusion is supported by the evidence here that RNA polymerase II is preferentially recruited to the risk alleles, at least in breast-cancer cell lines. Binding of RNA polymerase II was reported to mirror transcription levels,[23] but it is interesting to note that within close vicinity of rs2981578, there are multiple binding sites for transcription factors associated with transcriptional repression (NANOG, SIN3A, YY1, and HDAC2) rather than transcriptional activation. Our findings therefore suggest that the risk SNPs can affect transcriptional regulation but do not necessarily determine the direction of change.

We also examined the likely target gene of the 10q26 risk locus. Because of a lack of expressed SNPs in *FGFR2* in the heterozygous cell line MDA-MB-134, it was not possible to assay allele-specific transcription. In a surrogate analysis we found that *FGFR2* expression is decreased after si*FOXA1* transfection. Furthermore, we have recently found that FGFR2-regulated genes cluster near GWAS hits, further supporting a role for FGFR2 in mediating risk.[27] We attempted to assess associations between the newly identified potential causative SNPs and gene expression in 2,000 breast tumor samples from the METABRIC study but did not find evidence of association with expression of *FGFR2* or any of the neighboring genes, a result consistent with other recent studies.[32,33] However, our 3C studies confirm that the risk region interacts with the *FGFR2* promoter in both ER^+ and ER^- cell lines. These findings,

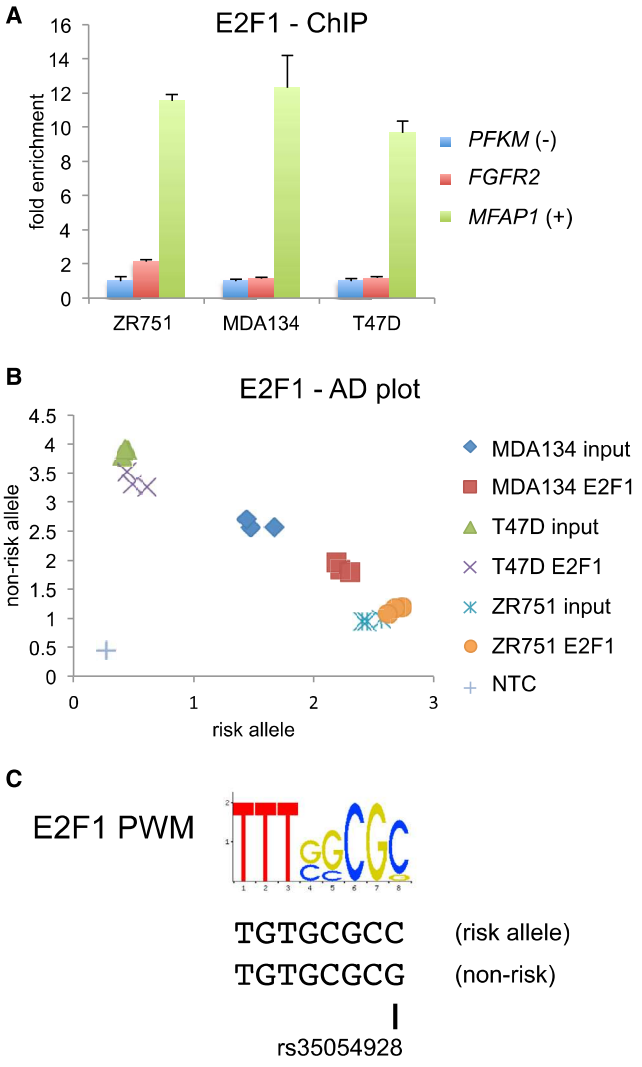

Figure 4. ChIP Assay for E2F1 at the *FGFR2* Risk SNPs in ZR-75-1, MDA-MB-134, and T47D Cell Lines

(A) Fold enrichment in a ChIP-qPCR experiment is shown relative to a negative control from the *PFKM* promoter (MIM 610681); *MFAP1*, positive control. Error bars show the standard deviation for three technical replicates of a representative experiment.

(B) Allelic discrimination plot via the rs2981578 TaqMan probe; NTC: nontemplate control.

(C) PWM of E2F1 from JASPAR, with the risk (C) and nonrisk (–) alleles of rs35054928 shown underneath.

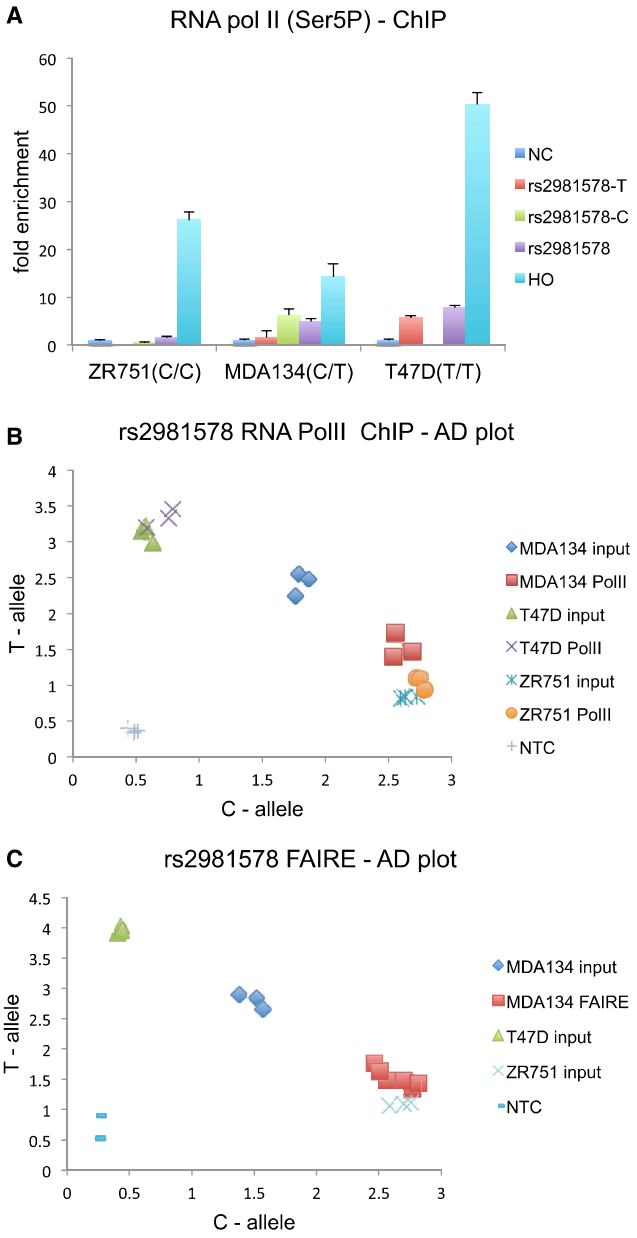

Figure 5. Allele-Specific Polymerase Binding and Chromatin Accessibility

(A) ChIP assay at the rs2981578 site with Ser5P-RNA Pol II-specific antibodies. Enrichment is given relative to a negative control (NC) primer pair from 8q24. The heme oxygenase promoter (HO) was used as positive control and error bars show the standard deviation of three technical replicates of a representative experiment.

(B) Allelic discrimination plot for the same experiment. NTC: non-template control.

(C) Allelic discrimination plot of DNA at rs2981578 isolated after FAIRE in the cell line MDA-MB-134. Genotyping results for T47D and ZR-75-1 cells are shown as controls.

together with the known critical role of FGFR2 in the developing mammary gland,[34] make *FGFR2* the most likely target gene for mediating risk. Although additional targets cannot be excluded, our conclusion is also consistent with the recent description of FGFR2 kinase activating mutations in breast cancer.[35] It is possible that the effect of regulatory SNPs can be detected only when examining expression at the correct developmental stage or after the relevant cell signaling stimuli.

The cell type in which *FGFR2* risk SNPs mediate their function has been debated.[10,12,13] Most expression correlations have employed whole-tissue sections. Ex vivo studies of skin fibroblasts have detected higher *FGFR2* expression levels in the presence of the tagging risk SNPs.[12] However, we note that in normal breast tissue, *FGFR2* is primarily expressed in epithelial cells.[13] Such an expression pattern mimics that of *FOXA1*, whose expression is also restricted to the breast epithelium. These observations are consistent with FOXA1 contributing to the risk phenotype by

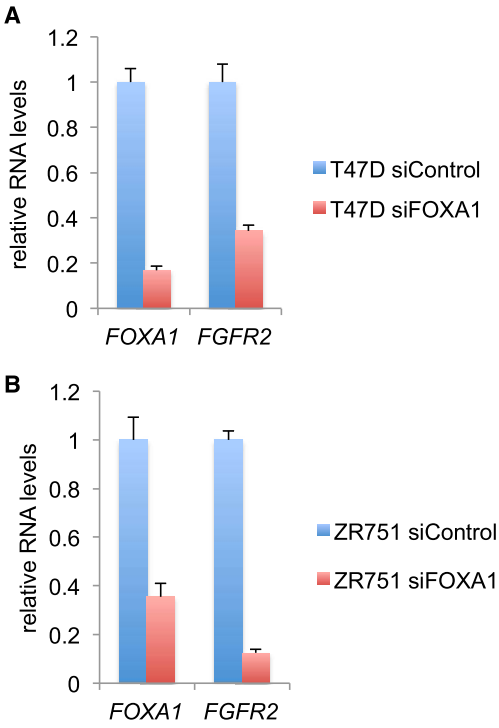

Figure 6. Effect of siRNA Downregulation of *FOXA1* on *FGFR2* Expression

FOXA1 and *FGFR2* RNA levels in (A) T47D and (B) ZR-75-1 cells. Error bars represent variation in three independent transfection experiments.

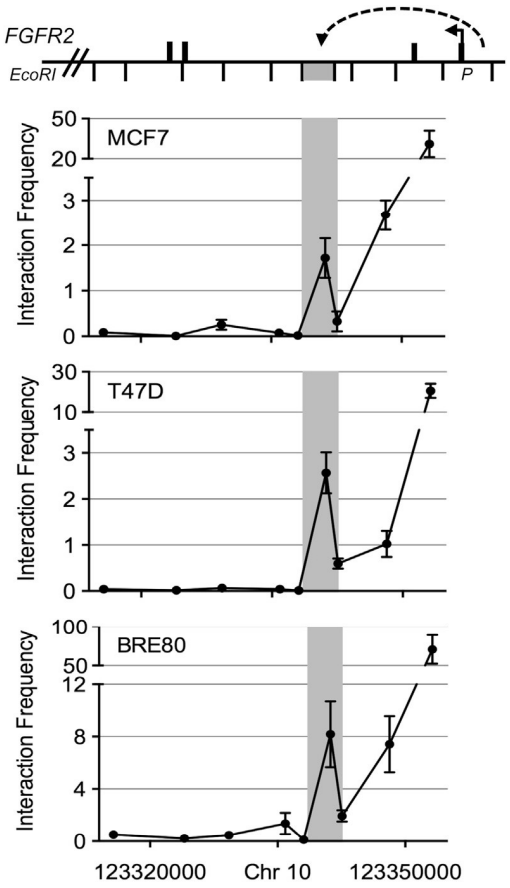

Figure 7. Chromatin Conformation Assays with the *FGFR2* Promoter as Bait

A diagram of the *FGFR2* locus indicating EcoRI sites is shown above the panels depicting normalized interaction frequencies in the cell lines indicated. MCF-7 and T47D are ER^+ breast cancer cell lines, and BRE-80 is an ER- transformed breast epithelial cell line. Error bars depict the standard deviation of three biological replicates assayed in duplicate. P, promoter; gray bar shows the risk region.

binding to rs2981578 in mammary epithelial cells, thereby promoting an increase in *FGFR2* expression. However, a more careful analysis of FGFR2 and FOXA1 will be required to determine colocalization of expression, especially during early mammary development when the risk SNPs might exert their function.

In conclusion, we have demonstrated that susceptibility to breast cancer at *FGFR2* is conferred by at least three independent signals, indicating the presence of at least three functionally relevant variants. We provide evidence that FOXA1 and E2F1 mediate risk, most likely exerting their effect in mammary epithelial cells.

Supplemental Data

Supplemental Data include Supplemental Acknowledgments, five figures, and eight tables and can be found with this article online at http://www.cell.com/AJHG/.

Consortia

The members of the GENICA network who contributed to this work are Hiltrud Brauch, Wing-Yee Lo, Christina Justenhoven, Ute Hamann, Yon-Dschun Ko, Christian Baisch, Hans-Peter Fischer, Thomas Brüning, Beate Pesch, Sylvia Rabstein, and Anne Lotz. For AOCS, see http://www.aocstudy.org/org_coll.asp; for kConFab, see http://www.kconfab.org/Organisation/Members.aspx.

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://browser.1000genomes.org
ENCODE, http://genome.ucsc.edu/ENCODE/
Human Genome Reference Consortium, http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
iCOGs, http://ccge.medschl.cam.ac.uk/research/consortia/icogs/
JASPAR, http://jaspar.genereg.net/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/
Primer3, http://bioinfo.ut.ee/primer3-0.4.0/primer3/
RefSeq, http://www.ncbi.nlm.nih.gov/RefSeq
Santa Cruz Biotech list of transcription factor binding sites, http://www.scbt.com/support-table-transcruz_gel_shift_oligonucleotides.html
UCSC Genome Browser, http://genome.ucsc.edu

References

1. Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R., et al.; SEARCH collaborators; kConFab; AOCS Management Group (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. Nature *447*, 1087–1093.
2. Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A., et al. (2007). A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat. Genet. *39*, 870–874.
3. Liang, J., Chen, P., Hu, Z., Zhou, X., Chen, L., Li, M., Wang, Y., Tang, J., Wang, H., and Shen, H. (2008). Genetic variants in fibroblast growth factor receptor 2 (FGFR2) contribute to susceptibility of breast cancer in Chinese women. Carcinogenesis *29*, 2341–2346.
4. Barnholtz-Sloan, J.S., Shetty, P.B., Guan, X., Nyante, S.J., Luo, J., Brennan, D.J., and Millikan, R.C. (2010). FGFR2 and other loci identified in genome-wide association studies are associated with breast cancer in African-American and younger women. Carcinogenesis *31*, 1417–1423.
5. Orr, N., Cooke, R., Jones, M., Fletcher, O., Dudbridge, F., Chilcott-Burns, S., Tomczyk, K., Broderick, P., Houlston, R., Ashworth, A., and Swerdlow, A. (2011). Genetic variants at chromosomes 2q35, 5p12, 6q25.1, 10q26.13, and 16q12.1 influence the risk of breast cancer in men. PLoS Genet. *7*, e1002290.
6. Michailidou, K., Hall, P., Gonzales-Neira, A., Ghoussaini, M., Dennis, J., Milne, R., Schmidt, M.K., Chang-Claude, J., Bojesen, S.E., Bolla, M.K., et al. (2013). Large-scale genotyping identifies 41 novel breast cancer susceptibility loci. Nat. Genet. *45*, 353–361.
7. Garcia-Closas, M., Hall, P., Nevanlinna, H., Pooley, K., Morrison, J., Richesson, D.A., Bojesen, S.E., Nordestgaard, B.G., Axelsson, C.K., Arias, J.I., et al.; Australian Ovarian Cancer Management Group; Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer (2008). Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet. *4*, e1000054.
8. Udler, M.S., Meyer, K.B., Pooley, K.A., Karlins, E., Struewing, J.P., Zhang, J., Doody, D.R., MacArthur, S., Tyrer, J., Pharoah, P.D., et al.; SEARCH Collaborators (2009). FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation. Hum. Mol. Genet. *18*, 1692–1703.
9. Chen, F., Chen, G.K., Millikan, R.C., John, E.M., Ambrosone, C.B., Bernstein, L., Zheng, W., Hu, J.J., Ziegler, R.G., Deming, S.L., et al. (2011). Fine-mapping of breast cancer susceptibility loci characterizes genetic risk in African Americans. Hum. Mol. Genet. *20*, 4491–4503.
10. Meyer, K.B., Maia, A.T., O'Reilly, M., Teschendorff, A.E., Chin, S.F., Caldas, C., and Ponder, B.A. (2008). Allele-specific upregulation of FGFR2 increases susceptibility to breast cancer. PLoS Biol. *6*, e108.
11. Zhu, X., Asa, S.L., and Ezzat, S. (2009). Histone-acetylated control of fibroblast growth factor receptor 2 intron 2 polymorphisms and isoform splicing in breast cancer. Mol. Endocrinol. *23*, 1397–1405.
12. Huijts, P.E., van Dongen, M., de Goeij, M.C., van Moolenbroek, A.J., Blanken, F., Vreeswijk, M.P., de Kruijf, E.M., Mesker, W.E., van Zwet, E.W., Tollenaar, R.A., et al. (2011). Allele-specific regulation of FGFR2 expression is cell type-dependent and may increase breast cancer risk through a paracrine stimulus involving FGF10. Breast Cancer Res. *13*, R72.
13. Sun, C., Olopade, O.I., and Di Rienzo, A. (2010). rs2981582 is associated with FGFR2 expression in normal breast. Cancer Genet. Cytogenet. *197*, 193–194.
14. Schmidt, D., Wilson, M.D., Spyrou, C., Brown, G.D., Hadfield, J., and Odom, D.T. (2009). ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. Methods *48*, 240–248.
15. Carroll, J.S., Meyer, C.A., Song, J., Li, W., Geistlinger, T.R., Eeckhoute, J., Brodsky, A.S., Keeton, E.K., Fertuck, K.C., Hall, G.F., et al. (2006). Genome-wide analysis of estrogen receptor binding sites. Nat. Genet. *38*, 1289–1297.
16. Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D., and Carroll, J.S. (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. Nat. Genet. *43*, 27–33.
17. Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R., and Lieb, J.D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res. *17*, 877–885.
18. French, J., Ghoussaini, M., Edwards, S., Meyer, K., Michailidou, K., Ahmed, S., Kahn, S., Maranian, M., O'Reilly, M., Hillman, K., et al. (2013). Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. Am. J. Hum. Genet. *92*, 489–503.
19. Robinson, J.L., Macarthur, S., Ross-Innes, C.S., Tilley, W.D., Neal, D.E., Mills, I.G., and Carroll, J.S. (2011). Androgen receptor driven transcription in molecular apocrine breast cancer is mediated by FoxA1. EMBO J. *30*, 3019–3027.
20. Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al.; METABRIC Group (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.
21. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.
22. Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389–393.
23. Odawara, J., Harada, A., Yoshimi, T., Maehara, K., Tachibana, T., Okada, S., Akashi, K., and Ohkawa, Y. (2011). The classification of mRNA expression levels by the phosphorylation state of RNAPII CTD based on a combined genome-wide approach. BMC Genomics *12*, 516.
24. Nakshatri, H., and Badve, S. (2009). FOXA1 in breast cancer. Expert Rev. Mol. Med. *11*, e8.
25. Bojesen, S.E., Pooley, K.A., Johnatty, S.E., Beesley, J., Michailidou, K., Tyrer, J.P., Edwards, S.L., Pickett, H.A., Shen, H.C., Smart, C.E., et al.; Australian Cancer Study; Australian Ovarian Cancer Study; Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer (kConFab); Gene Environment Interaction and Breast Cancer (GENICA); Swedish Breast Cancer Study (SWE-BRCA); Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON); Epidemiological study of BRCA1 & BRCA2 Mutation Carriers (EMBRACE); Genetic Modifiers of Cancer Risk in BRCA1/2

Mutation Carriers (GEMO) (2013). Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat. Genet. *45*, 371–384, e1–e2.

26. Cowper-Sal lari, R., Zhang, X., Wright, J.B., Bailey, S.D., Cole, M.D., Eeckhoute, J., Moore, J.H., and Lupien, M. (2012). Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. Nat. Genet. *44*, 1191–1198.
27. Fletcher, M.N.C., Mauro, C.A.A., Chin, S.-F., Caldas, C., Ponder, B.A.J., Markowetz, F., and Meyer, K.B. (2013). Master regulators of the FGFR2 response in breast cancer. Nat. Commun. *4*. Published online September 17, 2013. http://dx.doi.org/10.1038/ncomms3464.
28. Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.
29. Kong, S.L., Li, G., Loh, S.L., Sung, W.K., and Liu, E.T. (2011). Cellular reprogramming by the conjoint action of ERα, FOXA1, and GATA3 to a ligand-inducible growth state. Mol. Syst. Biol. *7*, 526.
30. Vuaroqueaux, V., Urban, P., Labuhn, M., Delorenzi, M., Wirapati, P., Benz, C.C., Flury, R., Dieterich, H., Spyratos, F., Eppenberger, U., and Eppenberger-Castori, S. (2007). Low E2F1 transcript levels are a strong determinant of favorable breast cancer outcome. Breast Cancer Res. *9*, R33.
31. Ferrari, N., McDonald, L., Morris, J., Cameron, E., and Blyth, K. (2013). RUNX2 in mammary gland development and breast cancer. J. Cell. Physiol. *228*, 1137–1142.
32. Riaz, M., Berns, E.M., Sieuwerts, A.M., Ruigrok-Ritstier, K., de Weerd, V., Groenewoud, A., Uitterlinden, A.G., Look, M.P., Klijn, J.G., Sleijfer, S., et al. (2012). Correlation of breast cancer susceptibility loci with patient characteristics, metastasis-free survival, and mRNA expression of the nearest genes. Breast Cancer Res. Treat. *133*, 843–851.
33. Li, Q., Seo, J.H., Stranger, B., McKenna, A., Pe'er, I., Laframboise, T., Brown, M., Tyekucheva, S., and Freedman, M.L. (2013). Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. Cell *152*, 633–641.
34. Lu, P., Ewald, A.J., Martin, G.R., and Werb, Z. (2008). Genetic mosaic analysis reveals FGF receptor 2 function in terminal end buds during mammary gland branching morphogenesis. Dev. Biol. *321*, 77–87.
35. Reintjes, N., Li, Y., Becker, A., Rohmann, E., Schmutzler, R., and Wollnik, B. (2013). Activating somatic FGFR2 mutations in breast cancer. PLoS ONE *8*, e60264.